Tilburg University

On the standardized empirical process

Einmahl, J.H.J.

Published in:
Statistica Neerlandica

Publication date:
1989

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Einmahl, J. H. J. (1989). On the standardized empirical process. *Statistica Neerlandica*, 43(3), 175-179.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

On the standardized empirical process

J.H.J. Einmahl

*Dept. of Med. Informatics and Statistics
University of Limburg
P.O. Box 616
6200 MD Maastricht
The Netherlands*

In this note a new proof of a theorem of CSAKI (1975, 1982) concerning the almost sure behaviour of the standardized empirical process is established.

Key Words & Phrases: inequalities, order statistics, strong limit theorem.

Let $X_1, X_2, \dots$ be a sequence of independent and identically distributed random variables with common continuous distribution function F . Define the empirical distribution function at stage n ($n \in \mathbb{N}$) by

$$F_n(t) = n^{-1} \sum_{i=1}^n 1_{(-\infty, t]}(X_i), \quad t \in \mathbb{R},$$

and the corresponding empirical process by

$$U_n(t) = n^{1/2}(F_n(t) - F(t)), \quad t \in \mathbb{R}.$$

It is well known (CHUNG, 1949; SMIRNOV, 1944) that for the supremum of the absolute value of U_n a law of the iterated logarithm holds, i.e. that

$$\limsup_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |U_n(t)| / (\log \log n)^{1/2} = 2^{-1/2} \text{ a.s.}$$

A process which is perhaps more natural to consider is the so called standardized empirical process defined by

$$V_n(t) = n^{1/2}(F_n(t) - F(t)) / (F(t)(1 - F(t)))^{1/2}, \quad t \in \mathbb{R},$$

since now for each $t \in \mathbb{R}$ the random variable $U_n(t)$ is divided by its standard deviation, which of course implies that

$$\text{Var}(V_n(t)) = 1, \quad \text{for all } t \in \mathbb{R}.$$

Although it is known (JAMES, 1975) that for any $0 \leq \delta < 1/2$

$$\begin{aligned} & \limsup_{n \rightarrow \infty} \sup_{t \in \mathbb{R}} |U_n(t)| / ((F(t)(1 - F(t)))^\delta (\log \log n)^{1/2}) \\ &= 2^{1/2} \sup_{t \in \mathbb{R}} (F(t)(1 - F(t)))^{1/2 - \delta} = 2^{1/2} \cdot (1/4)^{1/2 - \delta} \text{ a.s.,} \end{aligned}$$

Received December 1988, Revised March 1989.

the process V_n fails to satisfy a law of the iterated logarithm. To be more precise, it is shown in CsÁKI (1975, 1982) that no standardization depending on n is possible, such that the limsup equals a finite positive value. His result reads as follows.

THEOREM. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of positive numbers.

(I) If $\sum_{n=1}^{\infty} a_n = \infty$, then

$$\limsup_{n \rightarrow \infty} (na_n)^{1/2} \sup_{t \in \mathbf{R}} |V_n(t)| = \infty \text{ a.s.}$$

(II) If $\sum_{n=1}^{\infty} a_n < \infty$ and $\{na_n\}_{n=1}^{\infty}$ is non-increasing, then

$$\lim_{n \rightarrow \infty} (na_n)^{1/2} \sup_{t \in \mathbf{R}} |V_n(t)| = 0 \text{ a.s.}$$

The proof of part I of this theorem is short and easy, see e.g. GAENSSLER and STUTE (1979, p. 208). The proof of part II is very technical and uses a delicate technique, in combination with a modified version of the Borel-Cantelli lemma. Generalizations of this technique have proven to be an extremely valuable tool in the study of almost sure results for related empirical processes, see EINMAHL and MASON (1985, 1988), EINMAHL and VAN ZUIJLEN (1988) and EINMAHL, HAEUSLER and MASON (1988). On the other hand it was surprising that no "standard" proof of part II of this theorem existed, i.e. a proof using an appropriate exponential probability inequality in combination with a so called maximal inequality.

It is the purpose of this note to present such a standard proof, thus showing that this theorem and its proof fit into the general theory of empirical processes. A minor added condition we have to assume is that $\{n^c a_n\}_{n=1}^{\infty}$ is non-decreasing for some (large) $c > 1$. The proof is presented below; finally some remarks are made.

PROOF. Before we begin with the actual proof we make some preparations. First we note that because of the well known probability integral transform we only have to prove the result for uniformly-(0,1) distributed random variables. (In that situation we will denote the empirical process by $\bar{U}_n$ instead of U_n .) Moreover, note that because of symmetry considerations and the fact that $(1-t)^{-1/2}$ is bounded on $(0, 1/2]$ it even suffices to show that

$$\lim_{n \rightarrow \infty} (na_n)^{1/2} \sup_{0 < t \leq 1/2} |\bar{U}_n(t)|/t^{1/2} = 0 \text{ a.s.}$$

The observation $\{a_n\}_{n=1}^{\infty}$ that $\{Ma_n\}_{n=1}^{\infty}$ ($1 < M < \infty$ arbitrary) also satisfies the conditions if does, gives that we only have to prove that for some finite absolute constant L

$$\limsup_{n \rightarrow \infty} (na_n)^{1/2} \sup_{0 < t \leq 1/2} |\bar{U}_n(t)|/t^{1/2} \leq L \text{ a.s.}$$

Consider

$$(na_n)^{1/2} \sup_{0 < t \leq a_n} |\bar{U}_n(t)|/t^{1/2}.$$

It is well known (see Remark 3 below) that the conditions imply

$$P(\min_{1 \leq i \leq n} X_i \leq a_n \text{ i.o.}) = 0, \quad (1)$$

which of course yields

$$\begin{aligned} & \limsup_{n \rightarrow \infty} (na_n)^{1/2} \sup_{0 < t \leq a_n} |\bar{U}_n(t)|/t^{1/2} \\ &= \limsup_{n \rightarrow \infty} (na_n)^{1/2} \sup_{0 < t \leq a_n} n^{1/2} t/t^{1/2} \\ &= \limsup_{n \rightarrow \infty} na_n = 0 \text{ a.s.} \end{aligned}$$

Hence it finally remains to show

$$\limsup_{n \rightarrow \infty} (na_n)^{1/2} \sup_{a_n \leq t \leq 1/2} |\bar{U}_n(t)|/t^{1/2} \leq L \text{ a.s.} \quad (2)$$

Now it is time to state the aforementioned exponential probability inequality and maximal inequality.

INEQUALITY 1 (see Inequality 1 in EINMAHL and MASON, 1985). There exist absolute constants $c_1, c_2, c_3 \in (0, \infty)$, such that for any $0 < \alpha < 1$

$$\begin{aligned} & P(\sup_{\alpha \leq t \leq 1} |\bar{U}_n(t)|/t^{1/2} \geq \lambda) \\ & \leq c_1 \log(2/\alpha) \exp(-c_2 \lambda^2 \psi(c_3 \lambda/(n\alpha)^{1/2})), \quad \lambda \geq 0, \end{aligned}$$

where $\psi: [0, \infty) \rightarrow (0, \infty)$ satisfies $x\psi(x)/\log x \rightarrow 2$ as $x \rightarrow \infty$.

INEQUALITY 2 (see Inequality 2.10 in EINMAHL, 1987). Let $0 < \alpha \leq \beta \leq 1$, $0 < \epsilon < 1$ and write $n_k = [(1 + \epsilon/12)^k]$, $k \in \mathbb{N}$. Then we have for all $k \in \mathbb{N}$ and for $\lambda > (2/\epsilon)^{1/2}$

$$\begin{aligned} & P(\max_{n_k < n \leq n_{k+1}} \sup_{\alpha \leq t \leq \beta} |\bar{U}_n(t)|/t^{1/2} \geq \lambda) \\ & \leq 2P(\sup_{\alpha \leq t \leq \beta} |\bar{U}_{n_{k+1}}(t)|/t^{1/2} \geq (1 - \epsilon)\lambda). \end{aligned}$$

We now begin with the actual proof, which has been reduced to a proof of (2). Applying Inequality 2 (with $\epsilon = 1/2$) we see by the Borel-Cantelli lemma that it suffices to show

$$\sum_k P(\sup_{a_{n_{k+1}} \leq t \leq 1/2} |\bar{U}_{n_{k+1}}(t)|/t^{1/2} \geq L/(2(n_k a_{n_k})^{1/2})) < \infty.$$

Next we use Inequality 1 to obtain sufficiently sharp upper bounds for these probabilities, assuming that k is large enough. We have, if $c_3 L \geq 2$,

$$P(\sup_{a_{n_{k+1}} \leq t \leq 1/2} |\bar{U}_{n_{k+1}}(t)|/t^{1/2} \geq L/(2(n_k a_{n_k})^{1/2}))$$

$$\begin{aligned}
&\leq c_1 \log \frac{2}{a_{n_{k+1}}} \exp \left[-\frac{c_2 L^2}{4n_k a_{n_k}} \psi \left(\frac{c_3 L}{2(n_k a_{n_k})^{1/2} (n_{k+1} a_{n_{k+1}})^{1/2}} \right) \right] \\
&\leq c_1 \log \frac{2}{a_{n_{k+1}}} \exp \left[-\frac{c_2 L^2}{4n_k a_{n_k}} \frac{2(n_k a_{n_k})^{1/2} (n_{k+1} a_{n_{k+1}})^{1/2}}{c_3 L} \log \left(\frac{c_3 L}{2n_k a_{n_k}} \right) \right] \\
&\leq c_1 \left(\log \frac{2}{a_{n_{k+1}}} \right) (n_k a_{n_k})^{\frac{Lc_2}{2c_3} \left(\frac{n_{k+1} a_{n_{k+1}}}{n_k a_{n_k}} \right)^{1/2}}. \quad (3)
\end{aligned}$$

Observe that $n^c a_n \uparrow$ implies that $(n_{k+1} a_{n_{k+1}})^{1/2} / (n_k a_{n_k})^{1/2} \geq (23/25)^{(c-1)/2}$ for large k . This yields, if $(Lc_2/(2c_3))(23/25)^{(c-1)/2} \geq 2$, that the number on the right side of (3) is bounded from above by $c_1 (\log(2/a_{n_{k+1}})) (n_k a_{n_k})^2$.

So it remains to show that

$$\sum_k (\log(2/a_{n_{k+1}})) (n_k a_{n_k})^2 < \infty.$$

First note that $\sum_n a_n < \infty$ and $a_n \downarrow$ imply

$$\sum_k n_k a_{n_k} < \infty. \quad (4)$$

From $n^c a_n \uparrow$ we have $a_n \geq a_1/n^c$. Hence, writing $c_4 = \max(2/a_1, 1)$, we see that

$$\begin{aligned}
(\log(2/a_{n_{k+1}})) (n_k a_{n_k})^2 &\leq (\log(c_4 n_{k+1}^c)) (n_k a_{n_k})^2 \\
&\leq (\log c_4) (n_k a_{n_k})^2 + c (\log n_{k+1}) (n_k a_{n_k})^2. \quad (5)
\end{aligned}$$

Since (4) implies $\sum_k (n_k a_{n_k})^2 < \infty$, the first term on the right side of (5) is summable in k . Hence it suffices to prove that

$$\sum_k k (n_k a_{n_k})^2 < \infty. \quad (6)$$

Observe that (4) and $n_k a_{n_k} \downarrow$ imply $kn_k a_{n_k} \rightarrow 0$ ($k \rightarrow \infty$), yielding that $k(n_k a_{n_k})^2 \leq n_k a_{n_k}$ for large k . This observation and again (4) yield (6). This completes the proof. $\square$

REMARK 1. Generalization of the Theorem to the multivariate case (EINMAHL and MASON, 1985) and to the univariate truncated (at a fixed order statistic) case (EINMAHL, HAEUSLER and MASON, 1988) can also be proved in the above way.

REMARK 2. The proof of Inequality 1 hinges upon the following local fluctuation inequality for empirical processes.

INEQUALITY 3. (see e.g. EINMAHL, 1987; cf. also Lemma 2.4 in JAMES, 1975). Let $0 < t \leq 1/2$. Then for any $\epsilon \in (0, 1)$

$$P\left(\sup_{0 \leq s \leq t} |\bar{U}_n(s)| \geq \lambda\right) \leq C \exp \left[\frac{-(1-\epsilon)\lambda^2}{2t} \psi \left(\frac{\lambda}{n^{1/2}t} \right) \right], \lambda \geq 0,$$

where C only depends on ϵ .

This inequality has proven to be a very powerful tool in the study of empirical processes. Its proof is based on a maximal inequality for partial sums. In that respect there is some similarity between Inequalities 2 and 3.

REMARK 3. The proof of (1) is simple. Even $\sum_n a_n < \infty$ and $a_n \downarrow$ imply (1). Here is a short proof. Note that $a_n \downarrow 0$, since $\sum_n a_n < \infty$. But this implies

$$\{X_n \leq a_n \text{ i.o.}\} = \{\min_{1 \leq i \leq n} X_i \leq a_n \text{ i.o.}\}. \quad (7)$$

From the Borel-Cantelli lemma we have that $\sum_n a_n < \infty$ yields $P(X_n \leq a_n \text{ i.o.}) = 0$. This in combination with (7) proves (1).

REFERENCES

- CHUNG, K.L. (1949), An estimate concerning the Kolmogorov limit distribution, *Transactions American Mathematical Society* 67, 36-50.
- CSÁKI, E. (1975), Some notes on the law of the iterated logarithm for empirical distribution function, in: P. Révész (ed.), *Colloquia Mathematica Societatis János Bolyai 11*, North-Holland, Amsterdam, 47-58.
- CSÁKI, E. (1982), On the standardized empirical distribution function, in: B.V. Gnedenko, M.L. Puri and I. Vincze (eds.), *Colloquia Mathematica Societatis János Bolyai 32*, North-Holland, Amsterdam, 123-138.
- EINMAHL, J.H.J. (1987), *Multivariate empirical processes*, CWI Tract 32, Amsterdam.
- EINMAHL, J.H.J., E. HAEUSLER and D.M. MASON (1988), On the relationship between the almost sure stability of weighted empirical distributions and sums of order statistics, *Probability Theory and Related Fields* 79, 59-74.
- EINMAHL, J.H.J. and D.M. MASON (1985), Bounds for weighted multivariate empirical distribution functions, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 70, 563-571.
- EINMAHL, J.H.J. and D.M. MASON (1988), Laws of the iterated logarithm in the tails for weighted uniform empirical processes, *The Annals of Probability* 16, 126-141.
- EINMAHL, J.H.J. and M.C.A. VAN ZUIJLEN (1988), Strong bounds for weighted empirical distribution functions based on uniform spacings, *The Annals of Probability* 16, 108-125.
- GAENSSLER, P. and W. STUTE (1979), Empirical processes: a survey of results for independent and identically distributed random variables, *The Annals of Probability* 7, 193-243.
- JAMES, B.R. (1975), A functional law of the iterated logarithm for weighted empirical distributions, *The Annals of Probability* 3, 762-772.
- SMIRNOV, N.V. (1944), Approximate laws of distribution of random variables from empirical data, *Uspehi Matematiki Nauk* 10, 179-206 (in Russian).

Received December 1988, Revised March 1989.